Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

Spatial-temporal Prediction of Air Quality based on Recurrent Neural Networks

Xiaotong Sun
School of Information,
Renmin University of China,
Beijing, 100872, China
sun_xiao_tong@ruc.edu.cn

Wei Xu
School of Information,
Renmin University of China,
Beijing, 100872, China
weixu@ruc.edu.cn

Hongxun Jiang
School of Information,
Renmin University of China,
Beijing, 100872, China
jianghx@ruc.edu.cn

Abstract

To predict air quality (PM2.5 concentrations, et al), many parametric regression models have been developed, while deep learning algorithms are used less often. And few of them takes the air pollution emission or spatial information into consideration or predict them in hour scale. In this paper, we proposed a spatial-temporal GRU-based prediction framework incorporating ground pollution monitoring (GPM), factory emissions (FE), surface meteorology monitoring (SMM) variables to predict hourly PM2.5 concentrations. The dataset for empirical experiments was built based on air quality monitoring in Shenyang, China. Experimental results indicate that our method enables more accurate predictions than all baseline models and by applying the convolutional processing to the GPM and FE variables notable improvement can be achieved in prediction accuracy.

1. Introduction

PM 2.5 refers to airborne particles less than 2.5 μm in the aerodynamic diameter which has been linked to many adverse health impacts, including cardiovascular and respiratory morbidity [13]. Hence, PM 2.5 plays a crucial role in addressing many public health and environmental concerns. And obtaining accurate local PM 2.5 concentrations prediction and developing early warning system to provide air quality information towards the citizen have become an obvious and imperative need.

Shenyang is the capital of Liaoning Province and the largest city in the Northeast China. It is also an important heavy industry base in China that focuses on equipment manufacturing. In recent years, the air quality in Shenyang has improved gradually, but still at a nationally backward level, especially during the heating season in winter.

China's environmental supervisors have recently issued the country's most comprehensive and toughest plans to control and reduce air pollution by the year 2018, setting stricter limits on the levels of PM2.5 in some regions. Many cities have already announced their updated targets to control PM2.5 and the target for Shenyang is the number of days with good air quality to reach 270 days in 2018.

A series of work is needed to attain the air quality targets, including establishing monitoring networks for providing actual concentrations, developing emission inventory, calculating air quality indices, computer modeling for predicting future PM trends, and so on. In this area PM2.5 concentration forecasting has been of great significance. First, when a high PM concentration level occurs, accurate prediction can help notice residents to reduce regulatory outdoor activities in advance which may decrease pollution and mitigate adverse healthy impact [21]. Secondly, predicting the future trends in PM.5 concentration is an effective way to provide information for deciding whether an area will meet the PM2.5 standards and quantifying the amount of reductions needed to meet those standards.

In this paper, we propose an innovative spatial-temporal PM2.5 concentration prediction framework, in which the ground monitoring air quality data sets are selected to construct time panels. And we take the factory emission data and meteorological data as supplements. In this framework, we select gated-loop unit (GRU), which is an extension of recurrent neural networks (RNN), to implement mining of data features and hidden patterns. The experiment was conducted on relevant data sets collected from the air monitoring project in Shenyang city.

The rest of the paper is organized as follows. In Section 2, we review the related work on air quality prediction and discuss the main differences between our study and previous researches. In Section 3, we

URI: https://hdl.handle.net/10125/59567
ISBN: 978-0-9981331-2-6

introduce a framework for PM2.5 concentration prediction in which features extracted from multiple data sources combine for air quality emergency management. Section 4 introduces study area, data, evaluation metrics and experimental results. Finally, Section 5 summarizes the paper and make further research suggestions.

2. Literature review

High-quality air pollutant concentration prediction is an important basis of air pollution early warning and effective emergency management. The goal is to track and simulate the propagation process of pollutants by analyzing the generation mechanism of pollutants, transmission pathways, and historical data. Thus, we can predict the concentration of pollutants in a certain place at a certain time. In the past few decades, many researchers have devoted themselves to exploring trends in the concentration of pollutants. These existing prediction studies can be divided into two categories through modeling methods: numerical models and statistical models.

Numerous studies have used numerical models to study the main chemical reactions during pollutant movement, and then to predict the concentration of pollutants through the real-time emission data of pollutants. McKeen et al. [15] stablished seven air quality prediction models based on emission inventories and applied them to ozone and particulate concentration predictions in eastern Texas and adjacent states in the United States, and the collection of chemical and aerosol measurement data through aircraft helps diagnose the source of model bias. Shimadera et al. [24] used off-line prediction of community multi-scale air quality model (CMAQ) in East Asia and found that the extreme pollution of PM2.5 was mainly attributed to meteorological (wind power and wind direction) conditions rather than emission increase. However, they did not rule out the cause of large emissions created by winter heating in cold regions of China. Chuang et al. [4] used a new generation of regional air quality model (WRF/Chem) that uses on-line fully coupled meteorological models and chemical models to predict air quality in the southeastern United States and adds biogenic volatile organic compounds (BVOCs) variable. And adding emissions data of biogenic volatile organic compounds (BVOCs) effectively improves the accuracy of predictions. However, numerical models are computationally intensive models that are costly for routine prediction.

With the development of artificial intelligence and big data analysis, prediction models based on data mining and machine learning technologies are becoming more and more common. This type of model directly explores hidden patterns from the data. It does not require in-depth understanding of the physical and chemical properties of pollutants, and both computational efficiency and forecasting accuracy have been effectively improved. Commonly used machine learning methods include multiple linear regression (MLR) [2][17], Hidden Markov Model (HMM) [28], Geographic Weighted Regression (GWR) [13], Land Use Regression (LUR) [9], Support Vector Machine (SVM) [27], Neural Network Model [2][6][7] and so on. Mishra et al. [17] selected air pollutant data and meteorological data as independent variables, applied neural fuzzy model (NF) to develop a pollutant concentration prediction model based on historical data and weather conditions, and its predictions were better than those using MLR and artificial neural network (ANN) models. Ong et al. [19] introduced a deep recursive neural network (DRNN) to propose a new type of automatic encoder pre-training method for time series prediction tasks. Its prediction results are superior to the traditional training methods and it exceeds the forecast accuracy of the VENUS system currently being used by the Japanese government. However, these existing machine learning methods ignore the characteristics of the pollution sources, especially the pollutant data of the nearby pollutant discharge sites, and rarely consider the distances between the pollution sources and the monitoring points. In machine learning models, the absence of independent variables will have an adverse effect on the prediction results.

This paper presents a spatial-temporal prediction model of air quality based on GRU neural network, which is a spatial and temporal expansion of traditional air quality prediction model. GRU is an improved encoder-decoder model of the recurrent neural network (RNN) that performs well on many sequence learning problems. Compared with traditional RNNs and other common improved RNN models (such as long-term and short-term memory model), GRU has the advantages of simpler network structure, fewer parameters, and dealing with overfitting problems more effectively [3]. As GRU's computational efficiency and learning effects have gradually gained attention and recognition, they have begun to be applied in many different fields [10][29][31].

The main contributions of this study are the following. Firstly, estimating and predicting PM2.5 concentrations on an hourly time scale. Secondly, considering the impact of historical pollutants concentration, pollutants emission and meteorological condition on current and future PM2.5 concentrations. Thirdly, introducing GRU, a new deep learning technique, which is suitable for sequence learning modeling to PM2.5 concentrations forecasting. At last, apply real-time plant emissions data instead of emissions inventory to represent the effect of point pollution source on PM2.5 concentration.

3. Materials and methods

Numerous studies indicate that PM2.5 has a considerable negative impact on human health and more accurate PM2.5 concentration predictions can help mitigate its effects[18]. Considering that the formation and propagation process of PM2.5 is complicated and has correlations with environments, we proposed a spatiotemporal framework which incorporating ground pollutant measurement, factory emission, surface meteorological measurement, date and time data into the hourly PM2.5 concentration forecasting model. In particular, we introduced recurrent neural network model to capture the temporal autocorrelations of PM2.5 concentrations more effectively. The proposed framework consists of four main processes, namely data collection, data integration, data modeling, and spatiotemporal prediction (Figure 1).

3.1. Data Collection

Data collection is the first step in our framework during which we collected the data sets of variables required for PM2.5 concentration estimating. The selected data sets consist of monitoring data which are collected from related government departments and dummy data which are set to account for the monthly,

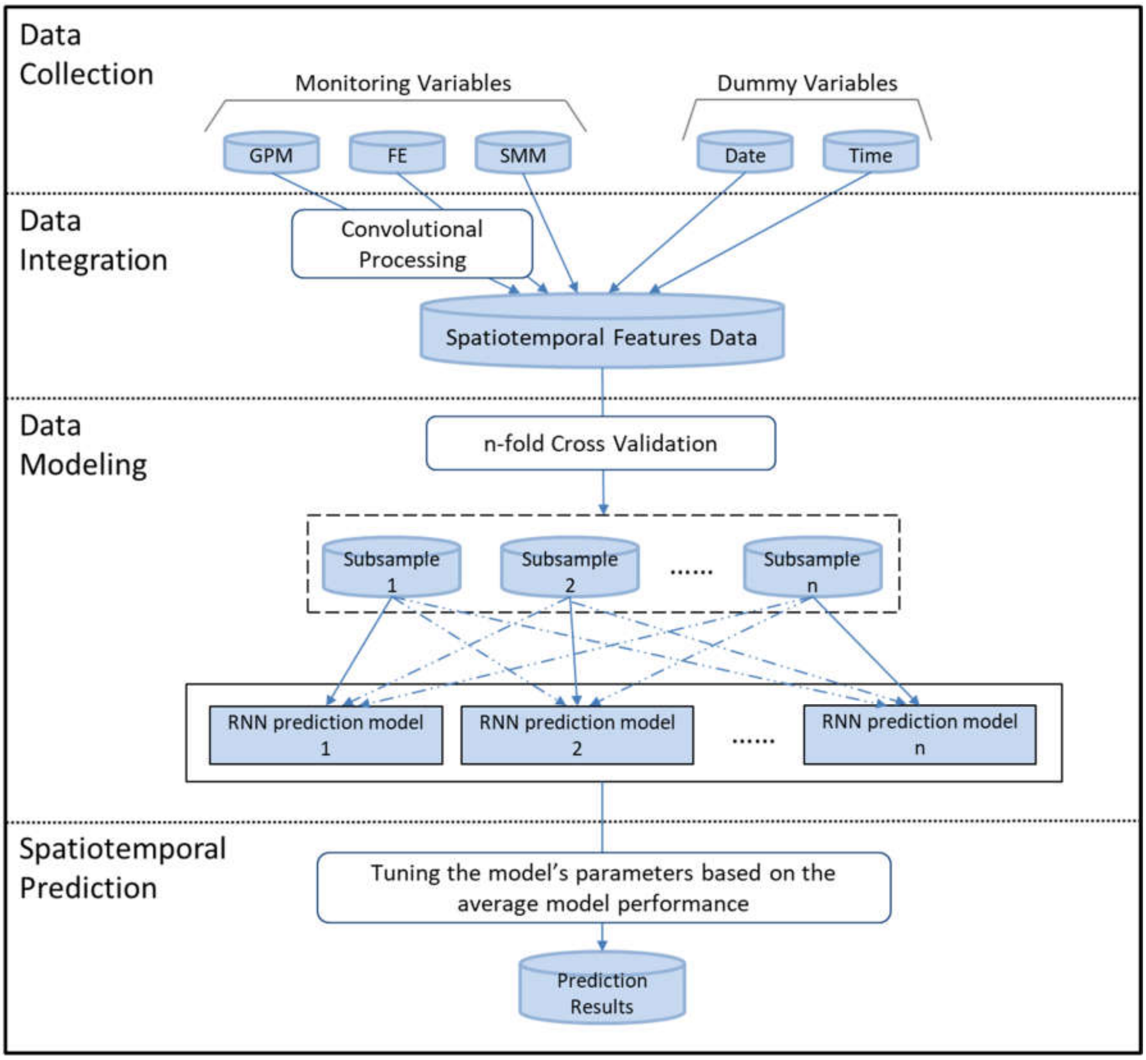

Figure 1. A framework for PM2.5 concentration forecasting

Table 1. Independent variables

Variable types		Variables	Data source
Monitoring	Ground Pollutants Measurement (GPM)	Hourly concentrations of CO, NO, NO_2, NO_X, SO_2, O_3, PM2.5, PM10	Related government departments
	Factory Emissions (FE)	Hourly emissions of particles, SO_2, NO_X (kg/h) Hourly benchmark gas flow (m^3/h)	
	Surface Meteorological Measurement (SMM)	Hourly atmospheric pressure (hpa), temperature (℃), humidity (%), wind direction (deg), wind speed (m/s)	
Dummy	Date	Month	
		Day	
	Time	Hour	

daily and hourly variation due to the human activities and meteorology.

3.1.1. Monitoring Variables. China established its ground monitoring network for air pollutants in late 2012 [11]. As the highest mean PM2.5 concentration always appears in winter and is believed mainly due to the heating, we took the ground pollutant measurement and factory emission data in winter from 2015 to 2017.

According to previous researches, PM2.5 concentrations are highly associated with the meteorological data and it is feasible to predict PM2.5 concentrations based on various surface meteorological measurement variables [2][4].

3.1.2. Dummy Variables. Atmospheric mechanism is complex and the relationship between monitoring variables and PM2.5 concentrations are time-dependent. In order to explore the implicit relationship better, we used monthly, daily and hourly dummy variables to illustrate the temporal variations.

The specific selections of variables are listed in Table 1.

3.2. Data Integration

3.2.1. Convolutional Processing. It is indisputable that local PM2.5 concentrations are affected by nearby point source pollutant emission and nearby PM2.5 concentrations and taking spatial information into consideration can improve model performance. We used convolutional processing incorporating these nearby variables to account for PM2.5 measurement and factory emission spatial correlations and PM2.5 measurement spatial autocorrelations. This convolutional processing was developed from the distance-inversed weighted average function proposed by Di et al [5]. We improved this process from two aspects.

First, unlike previous papers which only considered the influence of distances, we utilized the azimuth between nearby sites and local site to group nearby sites and combined wind direction data with the grouping results in the GRU model. The introduction of azimuth can help better simulate the spread of pollutants in the atmosphere and interpret the influence of nearby sites on local PM2.5 concentrations. We divided the emission factory sites or nearby pollutant monitor sites into eight groups according to the azimuths, taking 0, 45, 90, 135, 180, 225, 270 and 315 as the dividing point (Figure 2).

Second, we defined an attenuation coefficient λ for the process of pollutants passing through the atmosphere to describe the dissipation of pollutants from nearby sites to the local site. The coefficients are affected by distance, wind speed, humidity, air pressure, and temperature. For each group g of nearby site j, the kernel function can be expressed in general terms as

$$FE_{gj} = \sum_{i=1}^{n} \lambda_{ij} FE_i \tag{1}$$

$$GPM_{gj} = \frac{\sum_{i=1}^{n} w_{ij} GPM_i}{n} \tag{2}$$

$$\lambda_{ij} = e^{-\Lambda \frac{A_{ij} {d_{ij}}^2 h_j P_j}{v_j T_j}} \tag{3}$$

where FE_{gj} and GPM_{gj} are the values of convolutional layer, which represent pollutant levels at local site j in certain directions, and in each direction g there will be n nearby sites, FE_i is the emission of factory i and

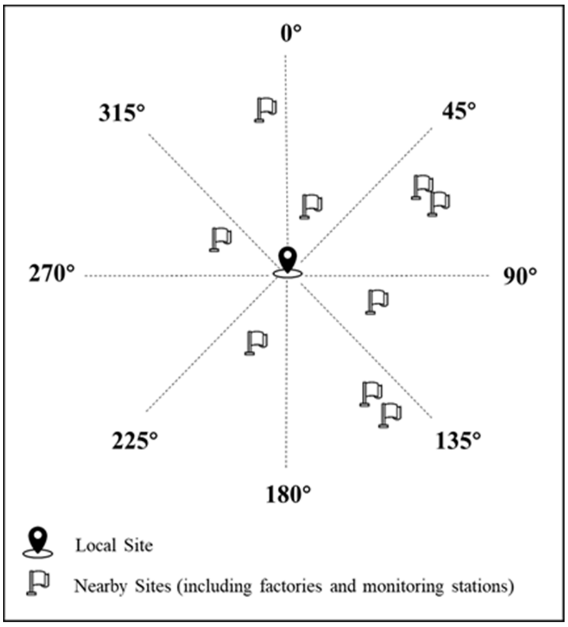

Figure 2. A sketch map of grouping sites by azimuths concentration forecasting

GPM_i is the pollutant concentration of nearby monitoring station i. λ_{ij} is the attenuation coefficient from nearby site i to local site j, and $w_{ij} \propto \lambda_{ij}$. Formula (3) is the function form of λ, where Λ is the attenuation constant, A_{ij} is the angle between azimuth of nearby site i and local site j and wind direction at local site j, d_{ij} is the distance between nearby site i and local site j, and h_j, v_j, P_j, T_j represent the humidity, wind speed, air pressure and temperature at local site j respectively.

We incorporated the convolutional processing results for nearby factory emissions and ground pollutant measurements with surface meteorological measurements as predictor variables to fit the recurrent neural networks.

3.2.2. Data fusion. Data fusion is a technology that makes full use of time and space information to integrate multi-source data together to obtain a unified representation, which improve data quality and help mine data value better.

We first conducted spatial data fusion on both original data sets and convolutional processing data sets based on longitude and latitude of monitoring sites. Among them, the original data sets include two categories: air quality hour data set and meteorological hour data set. And the convolutional processing data sets are consisted of air quality convolutional data set and factory emission convolutional data set. The air quality convolutional data set is obtained by convolving the hourly concentration of each pollutant in the air quality hour data set, while the factory emission convolutional dataset is obtained by convolving the hourly emissions of each pollutant in the factory emission hour data set.

Then the spatial-fused data were arranged in chronological order to prepare for the time-series data mining. Here we took monitoring variable time series length T (last T hours data) as the key parameter to improve the predictive capabilities. Let us take the case of T=3 as an example. At this time, the sample included the PM2.5 concentration at moment t, and the meteorological variable, ground monitoring variable, and convolution variable data at times t-1, t-2, and t-3.

3.2.3. Data preprocessing. In the data preprocessing stage, we performed four steps: data cleansing, data imputation, data transformation, and data standardization.

During the data imputation process, our fill values are calculated according to the following formula

$$Input_h = \alpha_1 avg_h + \alpha_2\left(\beta_1 V_{backw} + \beta_2 V_{forw}\right) \quad (4)$$

where $Input_h$ is the imputation of missing data, avg_h is the average of all the data at h o'clock, V_{backw} and V_{forw} are the most recent valid value before and after the missing value, β is the weight determined by the distance between the missing value and the nearest before and after valid value, and α is an experimentally determined weight that optimizes the filling effect. In our experiment, we took $\alpha_1 = \alpha_2 = 0.5$.

Data preprocessing is an indispensable step before these input data applied to training. It can help reduce the noise present in the data and improve the training performance. After data preprocessing, the required features have been extracted from the original data set.

3.3. Data Modeling

As discussed above, air quality prediction can be regarded as a certain type of time-sequence learning problem. Machine learning methods aim for more accurate prediction results by investigating and exploring hidden air pollution reacting and propagating patterns. And various environmental monitoring variable time series are applied in the prediction process. Given that the vanishing gradient problem prevents traditional RNNs from learning long-term dependencies, in this paper, we design a spatial-temporal model based on GRU, which were designed to combat the RNN problem through a gating mechanism but simpler than LSTM. In this model, we normalized the environmental variables to ensure that the proposed model can deal with features evenly. Moreover, in order to improve prediction accuracy and enhance procedural bias, techniques including 10-fold cross validation are applied to construct base GRU models.

When training the GRU model, back-propagation algorithm is applied to achieve the stochastic gradient descent according to the characteristics of the model structure. We reconstruct the dataset as dynamic time panel data and use the variable T to indicate the length of the time panel to see how the model's effect correlated with the time panel length. The training process continues until the weight matrix of the reaches convergence.

3.4. Spatiotemporal Prediction

After the training process, many GRU models are established to predict the future air quality. The final prediction model is selected by averaging calculations of several basic GRU predictors, which helps determine the dynamic time data panel length. We then make predictions in the condition of applying the determined length of time panel. The accuracy of predictions based on specific data set is compared to some baseline models to demonstrate the effectiveness of our proposed method.

4. The empirical analysis

4.1. Data description

Shenyang is an important central city in Northeast China and one of the national heavy industry bases focusing on equipment manufacturing. After entering the heating season every year, Shenyang, the capital of Liaoning Province, ushered in continuous and severe air pollution.

We define Shenyang as our study area and take ground pollutants measurements variables (GPM), factory emissions variables (FE), surface meteorological measurement variables (SMM) and temporal dummy variables into consideration.

4.1.1. Ground pollution measurement variables (GPM). Ground pollution measurement variables are used to reveal the PM2.5 concentration historical trends and the possible impact from other pollutants. We obtained ground pollutants measurement data from related government departments. Measurements data were collected from 11 monitoring sites and included hourly concentrations of eight pollutants (CO, NO, NO_2, NO_X, SO_2, O_3, PM2.5, PM10) from 2015 to 2017.

4.1.2. Factory emissions variables (FE). Factory emissions variables are used to reveal the level of pollutant emissions from factories which are considered as the main source of pollution in Shenyang. Factory emissions were also obtained from related government departments. Hourly emissions of particles, SO_2, NO_X and benchmark gas flow data were collected from 187 plants in Shenyang during winter from 2015 to 2017.

4.1.3. Surface meteorological measurement variables (SMM). Surface meteorological measurement variables is known to be a key factor that influences the formation and propagation process of PM2.5. Surface meteorological measurement were also obtained from related government departments. We selected hourly data of atmospheric pressure, temperature, humidity, wind direction and wind speed from 11 monitoring sites from 2015 to 2017.

4.1.4. Temporal dummy variables. The PM2.5 concentrations have been shown to exhibit some hourly variations. Hence, we include monthly, daily and hourly dummy variables to account for temporal variations.

4.2. Evaluation metrics

The 10-fold cross-validation (CV) technique is employed in the prediction results obtaining and validating process in our study. The entire training data

Table 2 Performances of various methods with all variables

Models	MAE	MSE	MAPE
MLR	8.7299	39.2241	21.54%
RF	6.5903	19.7879	10.44%
SVM	7.8910	21.3049	11.72%
ANN	8.0317	22.8971	12.85%
RNN	7.9118	22.9674	13.98%
LSTM	6.4810	18.0304	10.51%
GRU	4.6147	15.7878	6.29%

set was randomly divided into ten subsets, each subset containing approximately one tenth of the training data. In each round of cross-validation, we took nine subsets as training data and to make predictions for the remaining subset as testing data. The process was repeated 10 times until every subset was tested. Each trial will yield the correct rate (or error rate). The average value of the correct rate (or error rate) of the 10 times is used as an estimate of the accuracy of the algorithm.

When evaluating the model, we first arranged the data in chronological order and then take the last 30% as the hold-out test set. Model was trained on the rest 70% data and 10-fold CV was applied in this process. The averages of the CV evaluations verify the effectiveness of the model on the new dataset. And we can tune the model's parameters and complete model selection by comparing the average evaluations.

We calculated statistical indicators such as mean absolute error (MAE), mean absolute percentage error (MAPE) and root-mean-square error (RMSE) between predictions and observations in hold-out test set to assess the prediction accuracy of the proposed model for the entire study area and study period.

4.3. Experimental results

The experimental results under different parameter and variable settings are shown in this section.

4.3.1. Results of Model Validation. Previous studies have mainly applied multiple linear regression (MLR), random forest (RF), support vector machine (SVM) and artificial neural network (ANN) to investigate the hidden pattern of PM2.5 propagation. And as we take GRU, one of the traditional RNN models' extensions, as the base model in our framework, traditional RNN was also selected as baseline model in this comparative experiment. And as LSTM is the most popular version of RNN extension, we took LSTM as the benchmark model as well. All models run on our specific dataset under the framework we proposed. The result (Table 2) indicate that both GRU and LSTM have attained better prediction accuracy and our GRU-based prediction models are the most efficient.

4.3.2. Contribution of convolutional processing. We compared the evaluation scores between GRU-based models and all baseline models. The results of models with and without the data obtained by convolutional processing are shown in Figure 4.The results show that if we add variables which were produced by convoluting ground pollution monitoring data (Air quality convolutional variables) or factory emissions data (Emission convolutional variables) only, the performance will be slightly improved or even not be improved. And the effect of adding in emission convolutional variables were significantly better than the effect of adding in the air quality convolutional variables. However, when we add those two kinds of convolutional variables together, models show enhanced performances, especially in the GRU-based model we proposed with MAPE fallen rapidly from 53.74% to 10.18%, MAE fallen rapidly from 44.2176 to 4.6147 and RMSE fallen rapidly from 54.3399 to 6.2999. And overall, our proposed model performs best in all models.

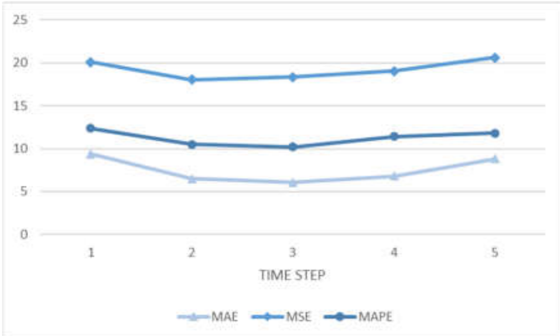

Figure 3. Impact of the length of the TIME panel

4.3.3. Evaluation of time-series modeling. The effects of dynamic time panel length T on the proposed GRU-based model is studied. The meaning of the key parameter T in the prediction framework has been described in Section 3.2.2. In the case of T = 1, 2, 3, 4, 5, the values of the evaluation metrics MAE, RMSE and MAPE of the proposed prediction model are obtained. The results are shown in Figure 3. As T increases, MAPE declines at the beginning and then slowly increases. The MAE curves and MSE curves have similar trends. When the length of the dynamic time panel is equal to 3, the proposed model achieves the best performance. The above experimental results show that the historical data of the hours before the predicted time does have effects on the PM2.5 concentration, because essentially RNN model learns the influence of the last-T-hour time series on the prediction target through the fixed time window. However, these performance metrics do not change significantly as the length of the dynamic panel changes.

4.4. Discussion

Experiments show that our proposed GRU-based framework for PM2.5 concentration prediction is effective. GRU-based deep learning models demonstrate their advantages in time series modeling, gradient disappearance problem solving, and model training efficiency. Based on the Shenyang City Air Pollution Monitoring Data from 2015 to 2017, we constructed a specific data set with relevant variables and PM2.5 concentrations. The comparative experimental groups conducted on this data set has revealed that the proposed learning framework promotes the improvement of PM2.5 prediction performance.

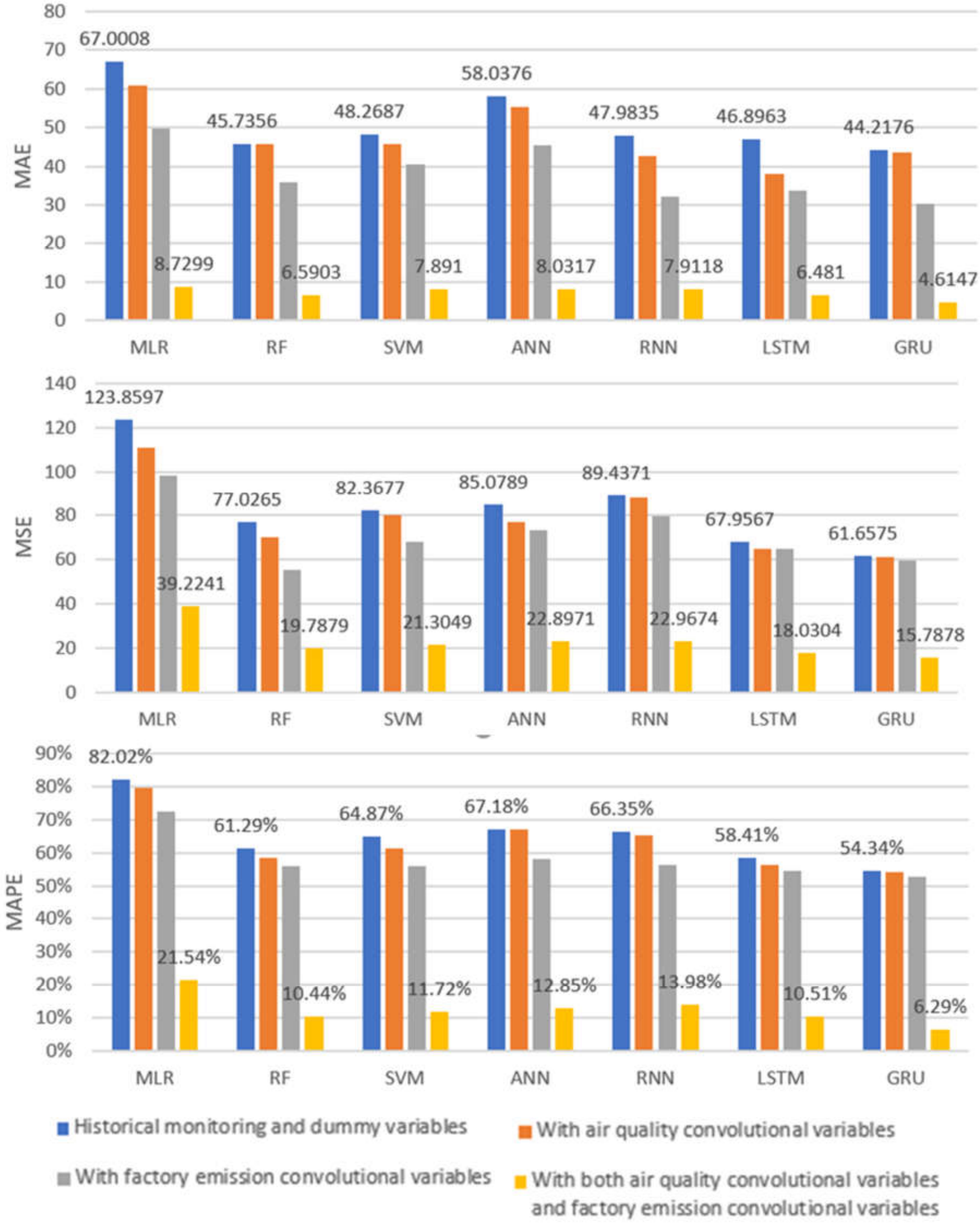

Figure 4. The performance of each model under different combinations of variables

Results of the model validation experiment indicate that GRU model is superior to traditional statistical methods and general RNN techniques in time series modeling.

Comparative experiments for the contribution of convolutional processing reveals that the inclusion of pollution emission information can improve the prediction capabilities more than the air quality convolutional variables which were the air pollutant concentration information at other monitoring sites. We think this is caused by the information loss in the process extracting original pollution information from the monitoring concentrations instead of using factory emissions directly. And when we add those two kinds of convolutional variables into the model in the meantime, the model may learn the propagation pattern from taking the hidden spatial information as a verification for the relationship between emissions and concentrations and use the pattern in prediction.

5. Conclusions and future work

In this paper, we propose a spatial-temporal GRU methods which is an extension of RNN in time and space to predict PM2.5. We use the characteristics of RNN to analyze the dependence of PM2.5 concentration time series. We design a convolution processing step to incorporate considerations of spatial information and pollution sources into the prediction process. The relevant experiments were carried out based on the data collected from the 2015-2017 Shenyang Air Pollution Project Monitoring Program. Empirical experiments validate the superiority of our proposed model and the significant improvements that convolutional variables can bring.

There are still some potential problems to be settled in future work. For example, we can divide 24 hours a day into several time intervals and analysis if there are differences between different intervals, for we need more accurate prediction during the time people are more likely to be outdoor. Applying our prediction framework to city groups, provinces, even nations is also meaningful work. And we can add remote sense data, like aerosol optical depth (AOD) data, into our framework, for these variables have been proved useful in PM2.5 estimation. It is believed that the performance of the PM2.5 prediction system for air quality emergency management can be improved to a new step with the inclusion of more comprehensive information.

References

[1] A. Zolghadri, and F. Cazaurang, “Adaptive nonlinear state-space modelling for the prediction of daily mean PM concentrations,” Environmental Modelling & Software, 2006, 21(6):885-894.

[2] Chen Y, Qin H, and Zhou Z G, “A comparative study on multi-regression analysis and BP neural network of PM2.5 index,” International Conference on Natural Computation. IEEE, 2014, 155-159.

[3] Cho K, Merrienboer B V, Gulcehre C, Bahdanau D, Bougares F, and Schwenk H, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” Computer Science, 2014.

[4] Chuang M T, Zhang Y, and Kang D, “Application of China in winter 2013 /Chem-MADRID for real-time air quality forecasting over the Southeastern United States,” Atmospheric Environment, 2011, 45(34):6241-6250.

[5] Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, and Schwartz J, “Assessing PM2.5 Exposures with High Spatio-Temporal Resolution across the Continental United States,” Environmental Science & Technology, 2016, 50(9):4712-4721.

[6] Feng X., Li Q, Zhu Y, Hou J, Jin L, and Wang J, “Artificial neural networks forecasting of pm 2.5, pollution using air mass trajectory based geographic model and wavelet transformation,” Atmospheric Environment, 2015, 107:118-128.

[7] J.B. Ordieres, E.P. Vergara, R.S. Capuz, and R.E. Salazar, “Neural network prediction model for fine particulate matter (PM2.5) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua),” Environmental Modelling & Software, 2005, 20(5):547-559.

[8] Hu X, Belle J H, Xia M, Wildani A, Waller L, and Strickland M, “Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach,” Environmental Science & Technology, 2017, 51(12):6936.

[9] Huang L, Zhang C, and Bi J, “Development of land use regression models for PM2.5, SO2, NO2 and O3 in Nanjing, China,” Environmental Research, 2017, 158:542-552.

[10] Kim P S, Lee D G, and Lee S W, “Discriminative Context Learning with Gated Recurrent Unit for Group Activity Recognition,” Pattern Recognition, 2017, 76.

[11] M Liu, J Bi, and Z Ma, “Visibility-based PM2.5 concentrations in China: 1957-1964 and 1973-2014,” Environmental Science & Technology, 2017, 51.

[12] M Popescu, SF Mihalache, and M Oprea, “Air Pollutants and Meteorological Parameters Influence on PM2. 5 Forecasting and Performance Assessment of the Developed Artificial Intelligence-Based Forecasting Model,” Revista de Chimie -Bucharest- Original Edition, , 2017, 68(4): 864-868.

[13] Madrigano J, Kloog I, Goldberg R, Coull B.A., Mittleman M. A., and Schwartz J, “Long-term Exposure to PM2.5 and Incidence of Acute Myocardial Infarction,” Environmental Health Perspectives, 2013, 121 (2), 192−196.

[14] Ma Z, Hu X, Huang L, Bi J, and Liu Y, “Estimating ground-level pm2.5 in china using satellite remote sensing,” Environmental Science & Technology, 2014, 48(13): 7436.

[15] Mckeen S, Grell G, Peckham S, Wilczak J, Djalalova I, and Hsie E. Y., "An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field study," Journal of Geophysical Research Atmospheres, 2009, 114(D7):D00F11.

[16] Mediavilla-Sahagún A, and Apsimon H. M., "Urban scale integrated assessment for London: Which emission reduction strategies are more effective in attaining prescribed PM air quality standards by 2005?," Environmental Modelling & Software, 2006, 21(4):501-513.

[17] Mishra D, Goyal P, and Upadhyay A, "Artificial intelligence based approach to forecast PM 2.5, during haze episodes: A case study of Delhi, India," Atmospheric Environment, 2015, 102:239-248.

[18] Neophytou A. M., Costello S. Brown, D. M., Picciotto S., Noth E. M., Hammond S. K., Cullen M. R., Eisen E, "Marginal structural models in occupational epidemiology: application in a study of ischemic heart disease incidence and PM2.5 in the US aluminum industry," American Journal of Epidemiology, 2014, 180(6):608.

[19] Ong B T, Sugiura K, and Zettsu K, "Dynamic pre-training of Deep Recurrent Neural Networks for predicting environmental monitoring data," IEEE International Conference on Big Data. IEEE, 2015, 760-765.

[20] Perez P, and Reyes J, "An integrated neural network model for PM10 forecasting," Atmospheric Environment, 2006, 40(16):2845-2851.

[21] Qin S, Liu F, Wang J, and Sun B, "Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models," Atmospheric Environment, 2014, 98:665-675.

[22] Reisen, Vald, rio Anselmo, Sarnaglia, Alessandro Jos, Queiroz, Reis Jr., Neyval Costa, and vy-Leduc, "Modeling and forecasting daily average PM10 concentrations by a seasonal long-memory model with volatility," Environmental Modelling and Software, 2014, 51(51):286-295.

[23] Reyes J M, and Serre M L, "An LUR/BME framework to estimate PM2.5 explained by on road mobile and stationary sources," Environmental Science & Technology, 2014, 48(3):1736-1744.

[24] Shimadera H, Hayami H, Ohara T, Yu M, Takami A, and Irei S, "Numerical Simulation of Extreme Air Pollution by Fine Particulate Matter in China in Winter 2013," Asian Journal of Atmospheric Environment, 2014, 8(1):25-34.

[25] Slini T, Kaprara A, Karatzas K, and Moussiopoulos N, "PM forecasting for Thessaloniki, Greece," Environmental Modelling & Software, 2006, 21(4):559-565.

[26] Song Y, Qin S, Qu J, and Liu F, "The forecasting research of early warning systems for atmospheric pollutants: a case in yangtze river delta region," Atmospheric Environment, 2015, 118(118), 58-69.

[27] Sun W, and Sun J. Daily PM2.5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm[J]. Journal of Environmental Management, 2016, 188:144.

[28] Sun W, Zhang H, Palazoglu A, Singh A, Zhang W, and Liu S, "Prediction of 24-hour-average PM(2.5) concentrations using a hidden Markov model with different emission distributions in Northern California," Science of the Total Environment, 2013, 443(3):93-103.

[29] Tang Y, Huang Y, Wu Z, Meng H, Xu M, and Cai L, "Question detection from acoustic features using recurrent neural network with gated recurrent unit," IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:6125-6129.

[30] Wang D, Wei S, Luo H, Yue C, and Grunder O, "A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine," Science of the Total Environment, 2017, 580:719-733.

[31] Wang Y, Liu M, Bao Z, and Zhang S, "Short-Term Load Forecasting with Multi-Source Data Using Gated Recurrent Unit Neural Networks", Energies, 2018.

[32] Xie Y, Wang Y, Zhang K, Dong W, Lv B, and Bai Y, "Daily Estimation of Ground-Level PM2.5 Concentrations over Beijing Using 3 km Resolution MODIS AOD," Environmental Science & Technology, 2015, 49(20):12280-12288.

[33] Xiya Zh, and Haibo H, "Improving Satellite-Driven PM2.5 Models with VIIRS Nighttime Light Data in the Beijing–Tianjin–Hebei Region, China," Remote Sensing, 2017, 9(9):908.

[34] Yang Y, and Christakos G, "Spatiotemporal Characterization of Ambient PM2.5 Concentrations in Shandong Province (China)," Environmental Science & Technology, 2015, 49(22):13431-13438.

[35] Yeganeh B, Hewson M G, Clifford S, Knibbs L. D., and Morawska L, "A satellite-based model for estimating PM 2.5, concentration in a sparsely populated environment using soft computing techniques," Environmental Modelling & Software, 2017, 88:84-92.

[36] Zhan Y, Luo Y, Deng X, Chen H, Grieneisen M. L, and Shen X, "Spatiotemporal prediction of continuous daily PM 2.5, concentrations across China using a spatially explicit machine learning algorithm," Atmospheric Environment, 2017, 155:129-139.

[37] Zong X, Zihan W, and Liu Y, "Optimization SVR fog prediction model based on genetic algorithm," Journal of Hebei University, 2016.